Exploring students learning approaches in MOOCs

Louis Faucon
Ecole Polytechnique FÃ©dÃ©rale de Lausanne
louis.faucon@epfl.ch

Prof. Pierre Dillenbourg (supervisor)
EPFL, CHILI

ABSTRACT

This study aims at understanding different students approaches for solving assignments in MOOCs. It makes use of a large dataset of logs from students interaction with the MOOC platform Coursera on a course about functional programming with Scala. In total more than 10.000 students participated in the assignments. Learning approaches are divided in two categories: starting with video lectures (**V**) and starting with the assignment (**A**); and students are divided in three groups: those applying purely the approach **V** , those applying purely the approach **A** and mixed-approach student who can apply both approaches.

We explore how our grouping correlates with assignment grades, number of submissions, time between submissions and overall performance. Significant difference has been found only on overall performance, while all three groups appear very similar on the other measures. Then we search correlations with approach changes for mixed-approach students. We observed that students are more likely to stay with the same approach, found significant difference on the starting time of learning activity sequences, but not on the time of student's first assignment submission. We found no correlation between the approach choice and the grade or number of submissions on the previous assignment.

INTRODUCTION

Massive Open Online Courses (MOOCs) attract every year thousands of students online which opens great opportunity for large scale analysis of their learning processes. The question of how do MOOC students learn is very difficult to answer as they have very different background, previous knowledge and motivation. This study explores the differences between learning approaches of thousands of students registered on an EPFL MOOC on the platform Coursera[1] about functional programming with Scala[2]. In particular, we are interested in the sequences of activities that a student performs when solving assignments. In MOOCs learning activities usually are one of Watching a Video, Reading or Posting on the Forum, Working on and Submitting Assignments and each student performs a sequence of these learning activities.

The normal expected approach to learning in MOOCs is to watch the video lectures, ask or answer questions in the forum concerning the difficult concepts of the lectures, then solve and submit assignments. However students are free to navigate through the content of the MOOC as they please when it is made available. This is why we expect to find navigation patterns diverging from the main learning approach. Bold pedagogical approaches let students face exercises before lectures to have them look for and build their knowledge and understanding by themselves, or simply to not lose the interest of students who already master the content of the lectures. This technique would be characterised in a MOOC by students trying to solve the assignments before watching the videos.

This project tries to answer the following research questions: **RQ1**: Do students differ from the main learning approach in MOOCs and try to solve assignment before watching the video lectures? **RQ2**: Does students overall performance correlates with their approach choices? **RQ3**: Do students change approach? **RQ4**: Do students approach choices correlate to their performance in the previous assignment or to the context in which they work on the current assignment?

PREVIOUS WORK

Navigation patterns

As MOOCs bring together learners from very different backgrounds, and allow free navigation through the content of the course, we can observe very different behaviour. This makes the study of MOOC students approaches and strategies a very broad field of research. For example, [4] and [3] show studies of students behaviour and interaction with MOOCs. [4] particularly explores the correlation between navigation strategies and demographics. They show that even students who complete the MOOC and earn the certificate skip about 22% of its content and that many students do not follow linearly the provided lectures. Also, [3] makes use of Markov models to simulate the behaviour of students and categorise them against their transition probabilities between learning activities.

Attrition and engagement

One of the main issue with MOOCs is the disengagement of students. Completion rates can sometimes be as low as 3% to 5% with averages not better than 10%. A lot of research has been conducted in order to understand and find ways to

[1]https://www.coursera.org/
[2]https://www.scala-lang.org/

reduce MOOC disengagement. For example [10] uses probabilistic models of student engagement, [5] classify engagement trajectories with the aim of developing features for particular groups of student and [2] explores students behaviour on MOOCs with a particular focus on engagement and attrition, categorising the students into three groups of *auditors*, *active* and *qualified* students.

Intelligent tutors and Bayesian Knowledge Tracing

The task of modelling student's knowledge and learning has been broadly studied in several fields such as Intelligent Tutors ([1] and [6]) or Bayesian Knowledge Tracing ([13] and [9]). Intelligent Tutors aim at providing artificial teachers inside learning environments to understand and help students. BKT uses probabilistic models to predict students knowledge and learning.

DATASET

We work on a dataset of logs describing student's interaction events with a MOOC hosted Coursera. The event are of three type: `Forum`, `Video` and `Assignment`. Each type of event has specific meta-data describing it. The list of available fields used in this study can be found on table 1. More data than displayed in the table is available in the logs. For example the logs also contain detailed interaction of students with lecture videos which has been studied by [12], [8], [7] and [11] for example.

Forum	Video	Assignment
StudentID	StudentID	StudentID
Timestamp	Timestamp	Timestamp
EventSubType	EventSubType	EventSubType
	OpenTime	OpenTime
	VideoID	ProblemID
		Grade
		HardCloseTime

Table 1. Schema of log data from the MOOC

Preprocessing

On our dataset we perform in order the following preprocessing steps:

- We remove events after the end of the course; As the content of the MOOC (videos and auto-graded assignments) stays available for a long time after the end of the course, we remove the late events in order to study the behavior of students in the context of a MOOC, with regular release of video and assignments with deadlines.
- We remove the unnecessary meta-data.
- As we study approaches to solve assignments, we remove students not working on assignments; This group of studnets is often called *Viewers* or *Auditors* in the MOOC research litterature [2].
- We remove the first assignment of the course as it is only used to set up the working environment for students and explain the submission process.

Extraction of learning patterns

In this study we do not consider the full sequence of event for each student, but divide these sequences in separated *patterns* for each assignment. We computed a matching from video to assignments based on the field `OpenTime`, then extracted from the students event sequences each sub-sequence corresponding to the work on a unique assignment. We are interested in exploring these shorter sub-sequences both between different students, but also between different assignment for each student.

After the preprocessing and the extraction the dataset contains over 40.000 learning patterns for about 10.000 students working on 6 assignments. Further more these learning patterns can be very long sequences as the data still counts about 10.000.000 recorded events.

LEARNING STRATEGIES

MOOCs are designed for students to, first, watch video lectures, then work on assignments, while reading and participating in the forum if they need or want. We want to explore learning approaches of students which diverge from the main approach of MOOCs. This is why this study will compare the following three groups of students:

- **Video First** (**V**): Students observed to always start working on assignment by watching the corresponding videos.
- **Assignment First** (**A**): Students observed to always start working on assignment by submitting.
- **Mixed approach** (**M**): Students observed to to use both type of approach

Figure 1 shows the counts of student for each group. As expected the main group is applying the approach **V**, which is the advised way of following the MOOC. The smallest group is composed of the students who start their learning sequences by working on assignments, certainly using a trial&error strategy. Finally we observed a significant group of students (**M**) that use both types of approach along the MOOC.

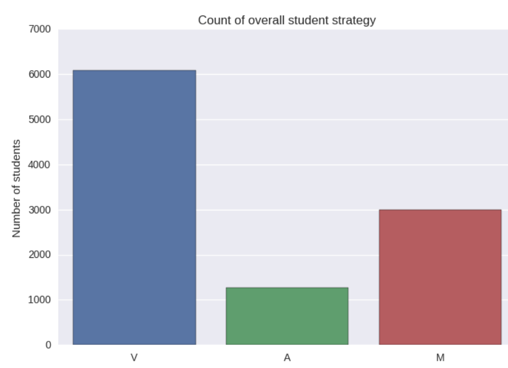

Figure 1. Counts of student in each group

Several reason could explain why student diverge from the approach **V**; Not having enough time before the deadline, already mastering the content of the course, higher motivation for assignments than watching videos, will of learning by doing. One of the promises of MOOC is to bring drastic changes from the conventional teaching and give more freedom to the learners. This divergence appear as a fulfilment.

Figure 2 shows the counts of student of each group submitting for each assignment. We can already observe that the groups **V** and **A** have a dropout rate of more than 50% along the whole MOOC while the group **M** has less than 30%. The next section explores more closely the differences between the three groups.

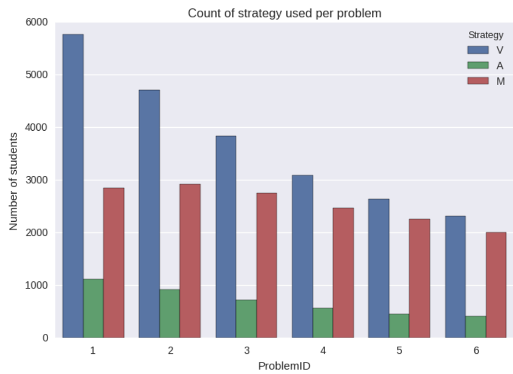

Figure 2. Number of student per group submitting for each assignment

ANALYSIS OF DIFFERENCES BETWEEN GROUPS V-A-M

In this section we analyse the different strategies and measure correlations between the three categories and other indicators of students behaviour such as their grades for the first and last submissions for each assignment, the average time they take before re-submitting a failed assignment, the average number of submissions and the overall final grade of the MOOC.

Assignment Grades

At first we want to compare the grades of the first assignment submission and the last submission for the different approaches. We expect: First, students with approach **V** would have higher grades on the first submission in average because they will apply the concept that they have just learnt; Secondly, students with approach **A** will have higher grades on their last submission if we consider that their approach choice shows us that they have higher motivation for the assignments. Even if our analysis shows a significant difference with small p-value (≤ 0.05) because of our very large number of students, the effect size is still very small (≤ 0.1). The detailed results are shown on table 2 and figure 3. This lead us to conclude that there is no meaningful difference between the three groups. The lack of variation on this measure is surprising. A possible explanation can be that students using approach **A** are right to chose not to watch the videos because they already know the concepts and are thus as able in average as the students watching the lectures.

	Group	**V**	**A**	**M**
First Problem	Average	8.85	8.98	8.89
	Effect Size	-0.05	0.09	0.01
Last Problem	Average	9.48	9.52	9.57
	Effect Size	-0.08	0.01	0.09

Table 2. Averages and effect size for assignment grades

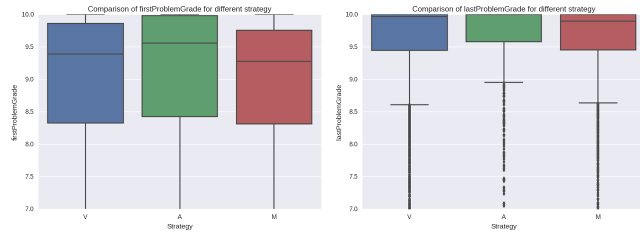

Figure 3. Assignment grades for first (left) and last (right) submission per group

Resubmission time

For similar reasons as on the grade comparison, we expect the different groups of students to behave differently in term of resubmission time. Indeed students that take the time to watch video (approach **V**) should be more likely to take more time before resubmitting an assignment after a wrong submission. However we again show that there are no meaningful variation between the three groups. Detailed measurements are given by figure 4 and table 3.

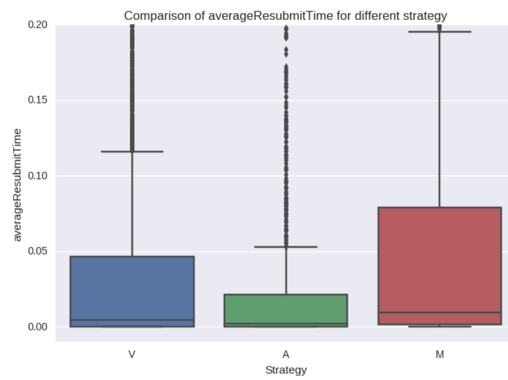

Figure 4. Average time between submission per group of student

	Group	**V**	**A**	**M**
Resubmission time	Average	0.14	0.15	0.17
(day)	Effect Size	-0.07	0.01	0.08

Table 3. Averages and effect size for resubmission time

Number of assignment submissions

As we expect differences in grades between the three learning approaches, we also expect variations in the number of submissions per assignment for each group. Figure 5 and table 4 show our measure of this feature. It reveals no strong difference between the three groups of students. As the need to submit several times and assignment shows a lack of understanding of the course content from the students, this result again support us to think that students deciding to skip the video lectures already master the topic and are as able as other students to successfully complete the assignments.

Final Grade

Even though no meaningful difference in performance of the three groups of students on single assignments was found, we can measure a significant variation in the students final grade. As the overall grade is computed as an average of

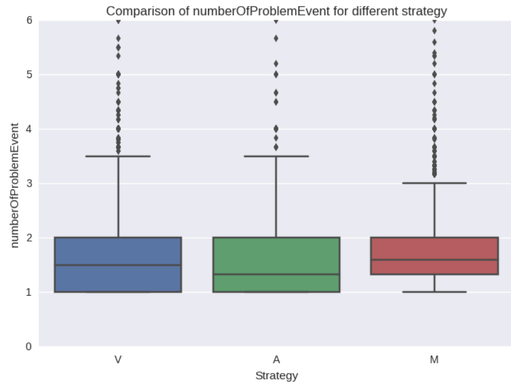

Figure 5. Average number of assignment submissions per group of student

	Group	**V**	**A**	**M**
Number of submissions	Average	1.72	1.66	1.75
	Effect Size	-0.003	-0.08	0.05

Table 4. Averages and effect size for number of assignment submission

the 6 assignments of the course, and considering that all the groups have similar grades, the difference is explained by the fact that students using approach **M** tend to stay longer in the course and solve more assignments. The results are shown in figure 6 and table 5. Mixed-approach students show a significantly higher final grade than the pure-approach groups. This result should be nuanced as doing more assignments will also lead to more chances to use both approaches for a student and thus be classified as a mixed-approach student.

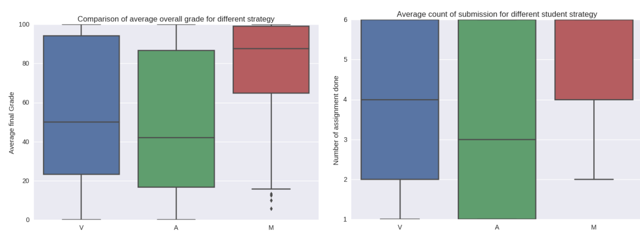

Figure 6. Overall grade (left) and count of distinct assignment submitted (right) per group of student

MIXED STRATEGY STUDENT

In this section, we analyse only the behaviour of the students observed to use the approach **M**. The aim is to understand what pushes students to decide whether to use approach **V** or **A** for an assignment. From figure 7 we can observe that the repartition of the mixed-approach students between the two choices of approach changes a lot along the 6 assignments of the MOOC. In particular, we observe that mixed-approach students watch more the videos in the beginning of the MOOC than at the end. This can be explained by the fact that by that time, students will be willing to do only the necessary work to pass the course and earn the certificate. We identify several possible causes for students to change learning approach that we can divide into two categories as displayed in table 6.

Time to deadline

	Group	**V**	**A**	**M**
Final Grade	Average	56.4	50.4	78.8
	Effect Size	-0.42	-0.40	0.71
Number of assignments done	Average	3.68	3.29	5.10
	Effect Size	-0.45	-0.44	0.75

Table 5. Averages and effect size for final grades and number of distinct assignment submitted

Category	Cause
Current context	- Time before deadline - Assignment difficulty
Previous context	- Grade - Number of submissions

Table 6. Possible causes of student changes of approach

A first reason for students to decide to try to do the assignment before watching the lecture videos could be that the student waited too long and the deadline for the assignment would force him to speed up his learning process. Our results shown on figure 8 show that there is a meaningful difference on the time at which students start their learning sequence, but their is no difference on the time at which they do their first submission. We can deduce from this that the students will decide to watch videos before doing the assignment if the deadline is further away, or that students who decide to use the approach **A** will procrastinate more and login to the MOOC only later.

Difficulty

An other reason for student to decide to use a different learning approach can be the difficulty of an assignment. Indeed we expect students to rely more on the lecture if the assignment is difficult than if they already know the content. As we do not have a measure of difficulty of assignment we decide to measure the difficulty using the average first grade, but only considering students that submitted for each of the 6 assignments. Figure 9 shows the average first grades. Comparing the grades with the the choice of approach given on figure 7

Our measures does not show any significant correlation between the difficulty and the approach choice. As we have only 6 assignments so we cannot reveal a trend. To answer this question a larger number of assignments is needed, thus the measure should be performed over several MOOCs, but the correlation could be hard to find as many other parameters could be involved.

Previous grade

For similar reasons as in previous section, we think that student grade on an assignment would be correlated to their choice of strategy for the next assignment. For example having a high grade will correspond to using the more *risky* approach **A**, while having a lower grade would push students to be more careful. Figure 10 shows the choice of approach of student depending on their grade. We identify no meaningful difference between the two choices, but show a significant correlation between lower grades and choice of Dropout.

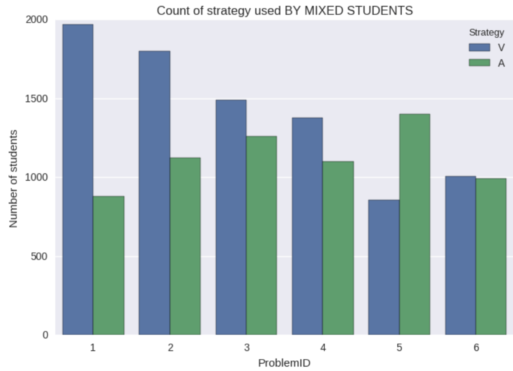

Figure 7. Choices of approach used by mixed-approach students for each assignment

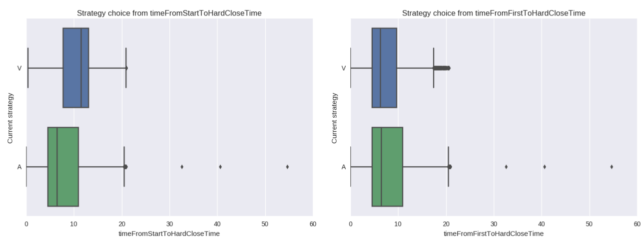

Figure 8. Choice of approach depending on the time from start of learning pattern (left) or first submission (right) to the deadline of the assignment

Previous number of submission

Figure 11 shows us the approach choice made from the number of submissions of students on the previous assignment. As a higher number of submissions shows that the student is experiencing difficulty with the assignment, we can think that students with a higher number of submissions will chose to use the approach **V**. No correlation has been discovered.

Previous approach

Figure 12 and table 7 show the probabilities of transition from the approaches **V** and **A** to either **V**, **A** or **D** (Dropout). We observe that, for each approach, student have a higher chance of repeating their choice than changing. This can be interpreted as due to the preferences of students who even if they use different approaches, still have a main learning approach. We observe that the approach **V** is more *stable* than the approach **A** and that students have a higher dropout probability after using the approach **A** than the approach **V**.

from \to	V	A	D
V	0.57	0.37	0.06
A	0.51	0.37	0.12

Table 7. Probabilities of transition from different approaches

CONCLUSION AND DISCUSSION

In this study we have extracted from students sequences of learning activities in a MOOC their learning approaches for each of the six assignments. We recognised two different approaches; starting with video lectures **V** or starting with the assignment **A**; and divided the students in three groups; those applying purely the approach **V** , those applying purely the

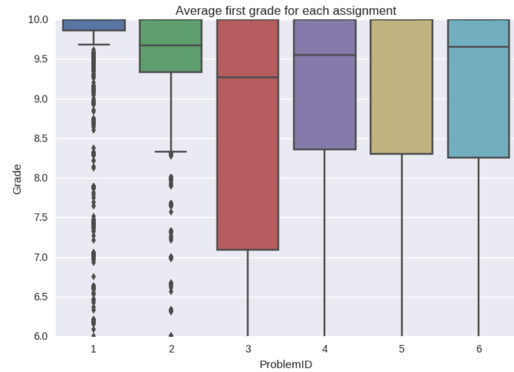

Figure 9. Grade of first submission from students having done every assignments

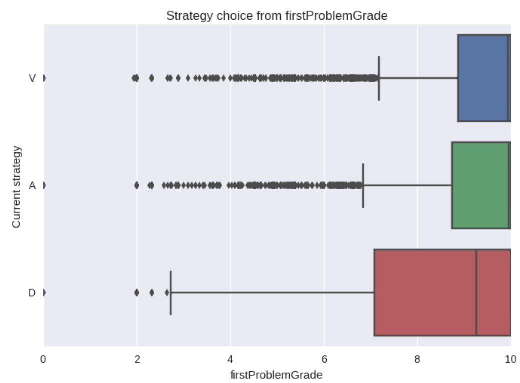

Figure 10. Choice of approach or Dropout on next assignment from previous first grade

approach **A** and those that apply both approaches, which we call mixed-approach student **M**. The fact that the groups **A** and **M** respectively contain about 10% and 30% of the students participating in assignments answers positively to **RQ1** and **RQ3**.

We first compared these different groups in terms assignment grade, number of submissions, time between submissions and overall performance. No meaningful variation was found on the assignment grades, number of submissions and time between submissions, which lead us to think that students choosing to use approach **A** are right that they do not need to watch the content of the videos and can obtain the same level of performance as other students. We found a significant difference between students observed to use both kind of approaches and pure-approach students which allow us to answer positively **RQ2**.

Then we explored the approach changes within the group of mixed-approach students. We found a significant difference in how long before the deadline students start their learning activity sequence depending on their approach, but no difference on the time of their first assignment submission; We had insufficient data to conclude on a correlation between the difficulty of assignments and the choice of approach; We found no correlation using the grade or number of submission on the previous assignment; Finally we observed that students

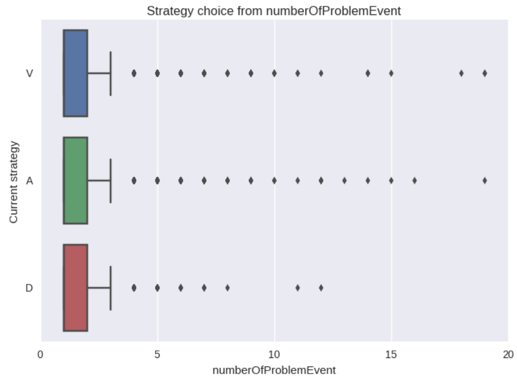

Figure 11. Choice of approach or Dropout from number of submissions on previous assignment

Figure 12. Choice of learning approach from previous approach

are more likely to keep the same approach from one assignment to the next. Thus, we cannot answer positively **RQ4** as our results have shown close to no difference on the observed parameters.

Our study would benefit of including several MOOCs in order to know if the proportion of students of each approach-group is similar between MOOCs of different topics and populations, if we find the same effects, or if some MOOCs are less easily accessible for students to use approach **A** efficiently. An other strong possible improvement is to not restrict our distinction of approaches on the first action and to explore more complex separations of students strategies which could include differences in video navigation or participation on the forum.

REFERENCES

1. Aleven, V., and Koedinger, K. R. Limitations of student control: Do students know when they need help? In *International Conference on Intelligent Tutoring Systems*, Springer (2000), 292–303.
2. Coffrin, C., Corrin, L., de Barba, P., and Kennedy, G. Visualizing patterns of student engagement and performance in moocs. In *Proceedings of the fourth international conference on learning analytics and knowledge*, ACM (2014), 83–92.
3. Faucon, L., Kidzinski, L., and Dillenbourg, P. Semi-markov model for simulating mooc students. In *Proceedings of the 9th International Conference on Educational Data Mining*, no. EPFL-CONF-218881 (2016).
4. Guo, P. J., and Reinecke, K. Demographic differences in how students navigate through moocs. In *Proceedings of the first ACM conference on Learning@ scale conference*, ACM (2014), 21–30.
5. Kizilcec, R. F., Piech, C., and Schneider, E. Deconstructing disengagement: analyzing learner subpopulations in massive open online courses. In *Proceedings of the third international conference on learning analytics and knowledge*, ACM (2013), 170–179.
6. Koedinger, K. R., Anderson, J. R., Hadley, W. H., and Mark, M. A. Intelligent tutoring goes to school in the big city.
7. Li, N., Kidzinski, L., Jermann, P., and Dillenbourg, P. How do in-video interactions reflect perceived video difficulty? In *Proceedings of the European MOOCs Stakeholder Summit 2015*, no. EPFL-CONF-207968, PAU Education (2015), 112–121.
8. Li, N., Kidziński, Ł., Jermann, P., and Dillenbourg, P. Mooc video interaction patterns: What do they tell us? In *Design for teaching and learning in a networked world*. Springer, 2015, 197–210.
9. Pardos, Z. A., and Heffernan, N. T. Modeling individualization in a bayesian networks implementation of knowledge tracing. In *International Conference on User Modeling, Adaptation, and Personalization*, Springer (2010), 255–266.
10. Ramesh, A., Goldwasser, D., Huang, B., Daumé III, H., and Getoor, L. Modeling learner engagement in moocs using probabilistic soft logic. In *NIPS Workshop on Data Driven Education*, vol. 21 (2013), 62.
11. Shi, C., Fu, S., Chen, Q., and Qu, H. Vismooc: Visualizing video clickstream data from massive open online courses. In *Visualization Symposium (PacificVis), 2015 IEEE Pacific*, IEEE (2015), 159–166.
12. Sinha, T., Jermann, P., Li, N., and Dillenbourg, P. Your click decides your fate: Inferring information processing and attrition behavior from mooc video clickstream interactions. *arXiv preprint arXiv:1407.7131* (2014).
13. Yudelson, M. V., Koedinger, K. R., and Gordon, G. J. Individualized bayesian knowledge tracing models. In *International Conference on Artificial Intelligence in Education*, Springer (2013), 171–180.